

8/pat3

10/537588

JC20 Rec'd PCT/PTO 06 JUN 2005

**Mixture of at least two fusion proteins as well as their production and use**

5

The present invention concerns a protein mixture comprising at least a first fusion protein comprising a protein or protein fragment, and an interaction domain and a protein translocation sequence, which effects that the fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially unfolded state and at least a  
10 second fusion protein comprising a protein or protein fragment, and an interaction domain and a protein translocation sequence which effects that the fusion protein is translocated through the cytoplasmic membrane upon expression in a bacterium in an essentially folded state, wherein the interaction domain of the first protein can bind to those of the second protein.

Phage display technology is currently used in many areas of biotechnology for  
15 identifying proteins with desired properties and enzymatic activities (Forrer, P. *et al.* (1999) Current Opinion in Struct. Biol. 9:514-520 and Gao, C. *et al.*, (2002) Proc. Natl. Acad. Sci. U.S.A. 99:12612-12616). Similarly, the technology is used to improve, for example, binding properties, the enzymatic properties and/or the thermodynamic stability of proteins already known or isolated by phage display technology (Forrer, P. *et al.* (1999) *supra*). The basis for  
20 the phage display technology lies in the observation that certain so called non-lytic bacteriophage merely infect bacteria and that the phage particles are not released by lysis of the bacterium but rather that the individual parts of the bacteriophage are transported through the cytoplasm into the periplasm and eventually to the bacterial cell surface where the complete phage is assembled which eventually disengages from the bacterial cell. The fusion  
25 of the protein of interest with a phage coat protein thus leads to the export of this protein from the bacterial cytoplasm and the presentation on the surface of the bacterium. Phage coat proteins suitable for presentation are for example pIII, pVI, pVII, pVIII and pIX derived from M13 phagemid (Gao, C. *et al.* (2002) *supra*).

The N-terminus of the phage coat protein is oriented towards the outside and,  
30 consequently, the fused protein has to be arranged N-terminally of the phage coat protein in order for it to be presented on the surface. This does not represent a problem, if single already known proteins are fused with one of the indicated phage coat proteins since the START and STOP codons of these proteins are known. It, however, leads to problems if a so called phage library has to be created wherein the phage coat proteins are fused with a cDNA library. The  
35 problem is caused by the fact that the coding nucleic acids comprised in the cDNA library

usually comprise translational STOP codons at the 3'-end since the cDNAs resulting from poly(A<sup>+</sup>) selection of the mRNA and from subsequent oligo-(dT)-priming always comprise translational STOP codons. Thus, a STOP codon will always be located between the cDNA and the phage coat protein upon fusion of an oligo-(dT)-primed cDNA 5' of the phage coat protein which in turn will inhibit expression of a fusion protein consisting of the cDNA encoded protein and the phage coat protein. Thus, Cramer, R. and Suter, M. (1993) *Gene* 137:69-75 developed a novel cloning and expression system based on the fact that the interaction domains of the two oncoproteins cJun and cFos were used, which form through a protein motive of regularly spaced leucine residues the so called "leucine zipper", a strong interaction between the two proteins (Landschulz *et al.* (1988) *Science* 240:1759-64) to connect the respective separately expressed phage coat protein and the cDNA encoded protein to form a heterodimer. For that purpose a fusion protein was expressed directed by a LacZ promoter which consisted of cJun and a C-terminus and of a phage coat protein (pIII) and, a second fusion protein which consisted of cFos at its N-terminus and of a cDNA library at its C-terminus, wherein also this protein was driven by a second LacZ promoter. Through the interaction between cJun and cFos via the respective leucine zipper within the periplasma of a bacterium the presentation of proteins and protein fragments, respectively, encoded by cDNAs became possible on filamentous phage.

When using the phage display technology there is the further problem that the assembly of the phage and, thus, the incorporation of the fusion proteins into the phage particles is carried out only in the periplasma (Russel *et al.* (1997) *Gene* 192(1):23-32). To export the respective fusion proteins into the periplasma of the bacterial cell an Sec signal sequence has to be added to the fusion protein by gene technological methods where applicable. This signal sequence causes the fusion protein to be transported in an essentially unfolded state into the periplasma. A large number of proteins, however, cannot be transported into the periplasma through the Sec transport pathway because the transport is inhibited by so called "stop-transfer" sequences or because of too rapid folding of the protein which occurs already in the cytoplasm. Stop-transfer sequences cause through the localized accumulation of positively charged amino acids in the protein sequence that the respective protein becomes stuck in the membrane upon translocation by the Sec transport pathway. Proteins which due to their rapid and/or stable folding cannot be bound in its unfolded form by proteins of the Sec transport pathway, in particular by SecB, are not transported through the Sec translocase complex and remain in the cytoplasm (Yamana *et al.* (1988) *J. Bio. Chem.* 263:19690-19696 and Berks, B.C. (1996) *Mol. Microbiol.* 22:393-404 and Bergs, B.C.

*et al.* (2000) *Mol. Microbiol.* 35:260-274). Proteins dependent on reducing conditions or which depend for their function on cytoplasmic co-factors like, for example, FeS centres or molybdopterin can also not reach the periplasma via the Sec transport pathway in functional form. Accordingly, many polypeptides due to the lack of compatibility with the Sec transport pathway cannot be presented in a functionally folded state by phage display and subsequently be selected. The translocation of fusion proteins through the Sec transport pathway into the periplasma, thus, represents a significant disadvantage of the phage display techniques known in the prior art.

From the different requirements on the cellular conditions for folding of certain proteins a further problem arises upon expression of fusion proteins, in particular in bacteria if one part of the fusion protein only attains a correct folding in the periplasma as is the case, for example, with antibody proteins (Gao, C. *et al.* (2002) *supra*) and the other part of the fusion protein can only be correctly folded in the cytoplasma as is the case, for example for green fluorescent proteins (GFP,) which is incompatible with Sec. Thus, the expression of, for example antibody-GFP-fusion proteins, i.e. fluorescently tagged antibody molecules is currently not possible in bacteria. The limitation to the Sec transport pathway, thus, prevents the production of a number of interesting protein conjugates, in particular in bacteria.

One object of the present invention is, thus, to overcome the limitation of the phage display technology of the prior art and to allow the production of fusion proteins which do not yield functional fusion proteins when produced by the prior art methods.

Thus, the present invention in one aspect provides a protein mixture comprising: a) at least a first fusion protein comprising: i) a protein or protein fragment, ii) an interaction domain and iii) a protein translocation sequence which effects that the fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially unfolded state and b) at least a second fusion protein comprising i) a protein or protein fragment, ii) an interaction domain and iii) a protein translocation sequence which effects that the fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially folded state, wherein the interaction domain of the first fusion protein can bind to those of the second fusion protein.

The protein or protein fragment of the first fusion protein comprises preferably proteins, which are translocated through the cytoplasmic membrane of the bacterium, preferably a Gram negative bacterium in an unfolded state and which accordingly do not require the reducing cytoplasmic environment and/or cytoplasmic co-factors for correct folding and which can also attain an essentially correct folding in periplasma. Examples of

such proteins comprise but are not limited to the immune globulin heavy chains, immune globulin light chains, fragments of these chains, so called "single-chain-antibody" (Bird, R.E. (1988) Science 242:423-6), diabodies (Holliger, P. (1993) Proc. Natl. Acad. Sci. U.S.A. 90(14):6444-8) receptors preferably extracellular domains of receptors like, for example, EGFR, PDGFR or VEGFR or receptor ligands like, for example, EGF, PDGF, or VEGF, integrines, preferably their extracellular domains, intimines and their domains, like for example EaeA, carbohydrate binding proteins and domains thereof like, for example, MBP and CBD, album binding proteins and domains or protein A and its domains.

The protein or protein fragment of the second fusion protein can be any protein or protein fragment preferred are, however, protein fragments which attain their folding and/or their function only if they are folded in the cytoplasm of a bacterium and which are thus translocated through the cytoplasmic membrane into the periplasma in an essentially folded state. Examples of such proteins are autofluorescent proteins like, for example, GFP or variants thereof with altered absorption maxima, enzymes like, for example,  $\beta$ -lactamase, co-factor dependent proteins like, for example, TMAO reductase and horseradish peroxidase, proteins which are encoded by a cDNA derived from a cDNA library or synthetic proteins.

In a preferred embodiment the protein or protein fragment of the first fusion protein and the protein translocation sequence is a phage coat protein or a periplasmatic marker enzyme, like PhoA, an intimin, a protein of the outer bacterial membrane or a periplasmatic receptor protein, in particular a carbohydrate binding protein. Preferred phage coat proteins which can be comprised in a protein mixture of the present invention are selected from M13 phagemid coat proteins pIII, pVI, pVII, pVIII and pIX. Out of these phage coat proteins only pIII and pVIII are provided with a known Sec dependent protein translocation sequence while the protein translocation sequences comprised in the remaining phage coat proteins have not been identified as of yet. Since these phage coat proteins are transported into the periplasma of the bacteria in an essentially unfolded state such proteins are considered as proteins which consist of a protein or protein fragment and a protein translocation sequence within the meaning of the invention without identification of the protein translocation sequence.

The interaction domains which are used in the first and the second fusion protein lead to binding of the first fusion protein to the second fusion protein. Thereby interaction domains are preferred which result in a relatively stable interaction between the two proteins, wherein a relatively stable interaction is an interaction which remains stable in the oxidative environment of the periplasma, on the bacterial cell surface or also outside the cell upon secretion of the heterodimer or heteromultimer. Suitable interaction domains of the first and

second fusion protein which can be comprised in the fusion protein according to the invention are, for example, a leucine zipper domain and a leucine zipper domain as they have been described for the first time in the two oncoproteins cJun and cFos (Landschulz *et al.* (1988) *supra*) or variants thereof derived from other hetero- or homodimers as well as artificial  
5 leucine zipper domains or helix-loop-helix-domains and helix-loop-helix-domains (Moor *et al.* (1989) *Cell* 56:777-783), a calmodulin and a calmodulin binding peptide (Montigiani, S. *et al.* (1996) *JMB* 258:6-13) or in each case of a peptide of a peptide dimer. The term interaction domain also comprises domains which allow the formation of multimers of more than two fusion proteins.

10 The protein translocation sequence of the first fusion protein effects that the fusion protein is translocated upon expression in a bacterium preferably in a Gram negative bacterium through the cytoplasmic membrane into the periplasma in an essentially unfolded state. Someone of skill in the art is capable of identifying suitable protein translocation sequences without undue burden by utilizing the following experiments. A protein sequence  
15 potentially suitable as protein translocation sequence, which leads to the translocation of a protein fused therewith in an essentially unfolded state, is used with a protein comprising a GFP-myc-TAG. If the potential protein translocation sequence does not lead to protein translocation into the periplasma the GFP protein is formed in the cytoplasm of the bacterium which can be detected via the cytoplasmic fluorescence. In this case it does not  
20 reach the surface or the media and, thus, the myc-TAG can neither be detected in the medium nor on the surface with an anti-myc-antibody, like for example the monoclonal antibody 9E10. If the sequence leads to translocation of the fusion protein into the periplasma and eventually to the presentation on the surface and secretion into the environment of the bacterium, respectively, the presented and secreted, respectively, GFP-myc-TAG fusion  
25 protein can be detected through an anti-myc-antibody in the medium and/or on the surface of the bacterium. At the same time no fluorescence should be detectable in the periplasma since upon translocation of the GFP into periplasma in an essentially unfolded state the protein will not be folded correctly (so called "Sec-incompatibility"). The protein translocation sequences which are preferably used in the first fusion protein are those which are recognized in the Sec  
30 dependent transport pathway (Danese, P.N. and Silhavy, T.J. (1998) *Annu. Rev. Genet.* 32:59-94) in the SRP dependent transport pathway (Meyer, D.I. *et al.* (1982) *Nature* 297:647-650) or in the YidC dependent transport pathway (Samuelson, J.C. *et al.* (2000) *Nature* 406:637-641). However, it can also be a transport pathway independent sequence. Particularly

suitable folding protein translocation sequences are, for example, signal sequences of PhoA, PelB, OmpA and pIII.

As a further element the second fusion protein comprises a protein translocation sequence which effects that the fusion protein is translocated through the cytoplasmic membrane upon expression in a bacterium, preferably in a Gram negative bacterium in an essentially folded state. A protein translocation sequence with this property is present, if a protein, for example, GFP which can only attain its functional confirmation in the cytoplasm of a bacterium, is transport into the periplasma without a loss of auto fluorescence. This property of the protein translocation sequence of the invention can be assessed with the experiment described above with respect to the first protein translocation sequence. With a similar experiment the consensus motive for the Tat specific leader peptide of the twin-arginine translocation (Tat) transport pathway of bacteria and plant chloroplasts have been determined. The Tat transport pathway known in the art allows the transport of proteins already folded in the cytoplasm into the periplasma and, thus, the transport of proteins into the periplasma which are incompatible with the Sec transport pathway. Similar to the transport through the Sec transport pathway also the Tat transport is mediated by a specific group of leader sequences (DeLisa, M.P. *et al.* (2002) J. Biol. Chem. 277:29825-29831). A further transport pathway known in the art which allows the transport of proteins in an essentially folded state is the one via thylakoid membranes (Settles, A.M. and Martienssen, R. (1998) Transcell Biol. 8:494-501). Accordingly, the second fusion protein comprises in a preferred embodiment of the present invention a signal sequence which is recognized by the Tat dependent transport pathway or by a thylakoid- $\Delta$ -ph dependent transport pathway and which, thus, leads to translocation of the fusion protein in an essentially folded state. A consensus motive of a protein translocation sequence recognized by the Tat dependent transport pathway is described in DeLisa, M.P. *et al.* ((2002) *supra*). The sequence is: S/T/RRXFLK.

In a preferred embodiment of the protein mixture of the present invention at least a first and at least a second fusion protein are covalently or non-covalently bound to each other. To attain a covalent bond between the two separately expressed fusion proteins it is possible to additionally place cystein residues or homologes thereof within the protein in the vicinity of the interaction domain, which will create a covalent bond between the two fusion proteins in the oxidative environment of the periplasma. Covalent bond can, for example, also be effected by the incorporation of amino acids with photoactivatable groups in both fusion proteins and subsequent UV-exposure of the proteins which are initially only bond to each

other non-covalently. Someone of skill in the art is aware of further methods to bind together to proteins, which are initially only bound together by non-covalent bonds. Methods known to a skilled person in order to covalently bind two fusion proteins which are non-covalently bound comprises, for example, psoralen crosslinking.

5 A further aspect of the present invention is a nucleic acid mixture which encodes a protein mixture of the present invention. A coding nucleic acid within the meaning of the present invention is a nucleic acid sequence which encodes a polypeptide of the invention or a precursor thereof. Preferably, the nucleic acid mixture is DNA or RNA, preferably a DNA, wherein the DNA can be single stranded or double stranded. The nucleic acid respectively  
10 encoding the first or the second fusion protein furthermore comprises promoters which allow the expression of the respective fusion proteins in the host cell. Suitable promoters for the expression in, for example, *E. coli*, are the trp promoter, lacZ promoter, tet promoter, T7 promoter or ara promoter. Further elements which can be present in the nucleic acids, which constitute the respective nucleic acid mixture, are origins of replication (Ori), selective marker  
15 genes which, for example, mediate ampicillin or chloramphenicol resistance. Aside from the region coding for the respective fusion proteins the nucleic acids can comprise those elements, which are usually employed in bacterial expression vectors. Someone of skill in the art is aware of a number of such elements as well as vectors like for example pGEM or pUC.

In a preferred embodiment of the nucleic acid mixture of the present invention the two  
20 nucleic acids coding for the first and the second fusion protein are covalently linked to each other, preferably via phosphor diester bond. In particular the nucleic acid molecules which code for the first and the second fusion protein and which comprise suitable regulatory elements are comprised on one plasmid, thus, allowing that the protein mixtures according to the invention can be prepared, for example, in a bacterium already by transfection of only one  
25 plasmid and by infection with only one phage, respectively, if the nucleic acid is comprised in a phage. In a preferred embodiment both fusion proteins are expressed under the control of only one promoter as bicistronic cassette.

A further aspect of the present invention is a vector comprising a protein mixture of the invention and/or comprising a nucleic acid mixture of the invention. A vector within the  
30 meaning of the invention is a protein-nucleic acid mixture, which is capable to introduce the protein mixtures and/or nucleic acid mixtures comprised therein into a cell. In that it is preferred that the fusion proteins encoded by the nucleic acid mixtures are expressed in the cells and that *de novo* synthesized fusion proteins can be recovered from the cells and can be

presented on the cell surface, respectively. Suitable vectors are, for example non-lytic phages, like M13 phage, fd phage, F1 phage and lytic phage, like  $\lambda$  phage.

A further aspect of the further invention is a cell comprising a protein mixture of the invention, a nucleic acid mixture of the invention and/or a vector of the invention. Cells of the invention can be prokaryotic or eukaryotic cells. In the preferred embodiment of the present invention the cells of the invention are prokaryotic cells, in particular bacteria and more preferably *E. coli* (TG1, XL-1, JM83, BL21) or *B. subtilis*.

A further aspect of the present invention is a library comprising at least two protein mixtures of the present invention, at least two vectors of the present invention and/or at least two cells of the present invention, wherein the proteins or protein fragments of the respective first or the respective second fusion protein are different from each other. Such a library can either comprise specifically selected different known proteins or protein fragments or the interaction domain and the protein translocation sequence on the first or the second, preferably the second fusion protein can be fused with a cDNA library, wherein the expression of these nucleic acids leads to a number of different first or second fusion proteins which respectively comprise different proteins or protein fragments. Preferably the cDNA part is expressed at the C-terminus of the fusion protein to thereby circumvent the previously described problem with N-terminal fusion of a cDNA. In a preferred embodiment the library comprises a large number of cells of the present invention when each cell produces a different protein mixture, preferably presents it on its surface. In case that the protein or protein fragment and interaction domain of the first protein is a phage co-protein the library of the present invention allows the presentation of a large number of proteins or protein fragments, which are comprised in the second fusion protein. The presentation is, thus, not limited as are the phage display libraries known in the prior art to proteins or protein fragments which fold into their functional form in the periplasma of the cell but also comprises proteins which can attain the functional folding in the cytoplasm.

The protein mixtures according to the invention which can form heterodimers or multimeres, wherein the components of the heterodimers or multimeres attain their three dimensional structure in at least two different cellular compartments can now be used in a number of methods comprising among others phage display.

A further aspect of the present invention is, thus, a method for identifying substances which can bind to a protein mixture, a vector of the present invention or to a cell of the present invention comprising the step:



- a) contacting at least one potential binding substance with a protein mixture of the invention, a vector of the invention or a cell of the invention and
- b) determining the binding of the substance to said protein mixture, said vector and/or said cell.

5           This method primarily serves the purpose of identifying a substance or substances, which can bind to an already known protein target, for example, to identify an inhibitor, an activator, competitor or modulator of the known protein target. The potentially binding substances the binding of which to a protein mixture of the invention, a vector of the invention and/or the cell of the invention should be measured can be any chemical substance  
10 or substance mixture. For example, it can be substances from a peptide library, substances from a combinatorial chemical library, cell extracts, in particular plant cell extracts and proteins or protein fragments.

          Contacting of the potentially binding substance(s) with a protein mixture, vector or cell of the invention is understood to mean any possibility of interaction between the two  
15 components wherein both components can be independently of each other in liquid phase, for example, in solution or in suspension, or can be attached to a solid phase, for example, to an essentially planar surface or can be in the form of particles, pearls or the like. In a preferred embodiment there is a plurality of different potentially binding substances immobilized on a solid surface and is contacted with the protein mixture of the invention, a vector of the  
20 invention or cells of the invention and subsequently binding of the substances of the invention to the various positions at which the respective different potentially binding substances are immobilized is measured.

          Measuring of binding of the protein mixtures, the vectors or the cells of the present invention to potentially binding substances can be carried out by measuring a marker  
25 connected to the protein mixture of the invention, the vector of the invention or the cell of the invention wherein suitable markers are known to the person skilled in the art and comprise, for example, fluorescence or radioactive markers. In a preferred embodiment the protein mixture, the vector, or the cell comprises in addition to the second fusion protein beside the protein or protein fragment the interaction of which with the potentially binding substance is to be investigated, an autofluorescent protein like, for example, GFP or variants thereof.  
30 Measuring the binding of the substance can also be detected via the change of electrochemical, in particular redox properties of, for example, the immobilized potentially binding substances after contacting. Suitable methods comprise, for example, potentiometric methods. Further methods for detecting the binding of two molecules or molecular mixtures

are known to someone of skill in the art and can all equally be employed for measuring the binding of the potentially binding substance to the protein mixture of the invention, the vector of invention or the cells of the invention.

If needed it is possible to introduce further steps prior to, in between or after the steps of the method of the invention like, for example, one or several washs after contacting to remove, for example, non-specific bonds between the potential binding substance and the protein mixture of the invention, the vector of the invention or the cell of the invention.

As a further step after measuring the binding of the substance the binding substance can be selected on the basis of, for example, the strength of the bond and can then be used directly, for example, for the inhibition of the known protein target. It is, however, also possible to modify the binding substance by methods known in the art which also comprise methods of combinatorial chemistry. For example, by adding halogen side groups, preferably F or Cl, by adding lower alkyl groups like methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl or tert-butyl groups or by adding amino, nitro, hydroxyl, amido, or carboxylic acid groups. The thus differently modified binding substances can then again be tested for the binding in the method of the invention and can be optimized with respect to the desired binding specificity and the effect caused thereby (for example, activation, inhibition or modulation of the respective activity).

A further aspect of the present invention is a method of identifying proteins or protein fragments, which bind to a test substance comprising the steps:

- a) contacting at least one test substance with a library of the present invention and
- b) measuring the respective binding of the test substance to the different protein mixtures, vectors and/or cells of the library of the present invention.

In this method protein or protein fragments are selected which can bind to a given test substance. Preferably those are proteins or protein fragments of the second fusion proteins, since this is correctly folded with a higher probability as compared to the proteins or protein fragments of the first fusion proteins which are only correctly folded, if the respective proteins can also attain their native conformation in the oxidative environment of the periplasm. A test substance within the meaning of the present invention can be any chemical substance or a mixture thereof. Preferably it is a protein or protein fragment, in particular a receptor or receptor ligand, a transcription factor, an ion channel, a molecule of the signal transduction cascade, a structure or storage protein, a toxin, a light receptor protein and pigment protein. Measuring of the respective binding of the various protein mixtures, vectors

and/or cells of the library to the test substance can be carried out as described above via marker dependent or marker independent assay methods.

In a preferred embodiment the method of the present invention comprises the further steps: Selecting at least one protein mix, one vector or one cell based on the measured binding  
5 and producing a second library wherein the library is produced by modification of the protein or protein fragment, which is comprised in the selected protein mix, in the selected vector or in the selected cell. The selection process of protein mixtures, vector or cells from the library is preferably carried out on the basis of the strength of the bond wherein protein mixtures, vectors or cells are preferred which show the strongest binding to the respective test  
10 substance. Starting from the amino acid sequence of the protein or protein fragment comprised in the selected protein mixture, vector or cell, which can be determined by standard methods, modification can be generated which respectively lead to minor changes in the amino acid sequence and thus to a multitude of derivatives which show a slightly different three dimensional structure in comparison to the starting protein and protein fragment, respectively.  
15 Such modifications can be obtained using methods known in the art like, for example, by random mutagenesis or by targeted substitution of single nucleic acid codons of the nucleic acid coding for the protein or protein fragment. It is thereby preferred that substitutions are so called "conservative" substitutions. A conservative substitution is present if, for example, a nucleic acid codon coding for a basic amino acid is replaced by another nucleic acid codon  
20 coding for a basic amino acid, a nucleic acid codon coding for another acidic amino acid is replaced by a nucleic acid codon, coding for a acidic amino acid and a nucleic acid codon coding for a polar amino acid is replaced by another nucleic acid codon coding for a polar amino acid, respectively.

The second library newly generated on the basis of the selected protein mixtures, vectors or cells can now again be contacted in a further step with the test substance  
25 whereupon in a further step the respective binding of the test substance to the modified protein mixtures, vectors or cells of the second library is measured. As the case may be it is now possible to repeat the steps of selecting at least one protein mixture, at least one vector or at least one cell on the basis of the measured binding and the subsequent production of a third and n-fold, respectively library as well as the contacting and measuring of the respective  
30 binding of the test substance to the various protein mixtures, vectors or cells of the third and n-fold library for one to n-fold times until a protein mixture, a vector or cell is selected which shows the desired binding.

The previously described method is also termed directed evolution since in a multitude of steps, which consist of modification and selection, proteins or protein fragments are further developed with respect to particular property in particular the binding property in an "evolutionary" way.

5           The proteins or protein fragments which have been identified or additionally have been optimized with respect to a particular property by above method can now be used as an active agent in a medicament, if they have been, for example, optimized for activation or repression of a particular cellular signal pathway. The same applies to binding substances which have been identified in methods for determining potentially binding substances. Thus, 10 the methods of the present invention comprise in a preferred embodiment the further step that the selected binding substance or the protein or protein fragment or a variant thereof comprised in the selected protein mix, in the selected vector or in the selected cell is admixed with a pharmaceutical acceptable carrier and/or auxiliary substance.

15           A "variant" of the protein or protein fragment comprises modifications of the N- or C-terminal or modification of amino acid side chains which, for example, increase the stability, solubility or biocompatibility of the proteins or protein fragments. Also comprised are fusion proteins of the proteins or protein fragments identified according to the invention which can comprise as a further component autofluorescent markers like, for example, GFP or cytostatic drugs like, for example, cholera toxin.

20           Pharmaceutically acceptable carriers and/or auxiliary substances comprises substances which stabilize the binding substance and the protein or protein fragments, respectively, or variants thereof, which increase the pharmaceutical tolerance or which are required by the respective form or application like for example tablet, band aid or infusion solution as, for example, preservative, buffer, salt or protease inhibitors.

25           A further aspect of the present invention is a kit for producing a mixture of nucleic acids according to claim 10 comprising:

- a) at least one first nucleic acid, comprising at least one restriction site 5' and/or 3' of a nucleic acid coding for a first fusion protein comprising:
  - 30           i) an interaction domain and
  - ii) a protein translocation sequence which effects that the first fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially folded state.

This kit allows the insertion of a chosen nucleic acid sequence 5' or 3' of the nucleic acid which codes for the interaction domain and the protein translocation sequence with the

result that the resulting nucleic acid codes for a fusion protein which comprises at its C-terminus and/or at its N-terminus a protein or protein fragment encoded by the respectively introduced nucleic acid sequence. Preferably, the introduced DNA is a cDNA library, wherein this is particularly preferred if it and has been introduced into the nucleic acid by using the 3'-restriction site. In a preferred embodiment the kit comprises the leucine zipper of the cFos protein and in a further preferred embodiment the Tat dependent protein translocation sequence TorA.

In a further embodiment of the kit the kit according to the present invention further comprises at least a second nucleic acid comprising at least one restriction site 5' and/or 3' of a nucleic acid coding for a second fusion protein comprising:

- i) an interaction domain and
- ii) a protein translocation sequence which effects that the second fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially unfolded state, wherein the interaction domain of the first fusion protein can bind to those of the second fusion protein.

This nucleic acid allows insertion 5' or 3' of the nucleic acid encoding for the interaction domain and the protein translocation sequence so that in result the resulting nucleic acid codes for a fusion protein which comprises at is N- or C-terminus a protein or protein fragment coded for by the inserted nucleic acid. For example nucleic acids coding for a phage coat protein can be inserted into a nucleic acid wherein those are preferably inserted at the 3' restriction site.

It has been shown that if nucleic acids coding for phage coat proteins are introduced into the second nucleic acid that the resulting fusion protein upon strong expression of, for example, the gIIIp-fusion protein lead to high toxicity in *E. coli* cells. For this reason an amber codon is inserted in classical phase display systems 5' of the gIII-protein. In suppressor strains (e.g. CL-1 Blue) the expression of the gIIIp-fusion protein is thereby reduced by 90 %. Furthermore the amber codon (which is read in non-suppressor strains as STOP codon) enables the easy soluble expression of the protein which was previously fused with a phage protein and presented on the phage by introducing the phagemid into a non suppressor strain (e.g. BL21) and expressing it therein. Accordingly, the first and/or the second nucleic acid comprises in a preferred embodiment either 5' or 3' an amber codon. Preferably, the amber codon is positioned in the first nucleic acid 5' and in the second nucleic acid 3'. Thereby it is possible that only the protein or protein fragment, which has been inserted into the first

nucleic acid 5' is expressed in a suitable host and at the same time that the toxic effect of the gIIIp which is inserted into the second nucleic acid 3' is prevented.

In a further preferred embodiment of the kit according to the invention the interactive domain of the second fusion protein is a leucin zipper domain of the cJun protein. In a further preferred embodiment the nucleic acid comprises a nucleic acid which codes for a Sec-dependent protein translocation sequence in particular the PelB leader peptide.

A further aspect of the present invention is the use of a cell for the production of a protein mix according to the invention as well as the use of a protein mix according to the invention, a vector according to the invention or a cell according to the invention for the preparation of a library according to the invention.

A preferred area of using the protein mixes of the invention, the phages of the invention, the cells of the invention in particular the libraries of the invention comprising the above referenced mixtures of proteins, phages and cells as well as of using the kits of the present invention is the presentation of proteins on filamentous phages. A particular focus thereby is on proteins which due to the incompatibility with the Sec transport pathway cannot be presented using the classical phage display technology. As a result of this presentation and selection of cDNA expression libraries and the presentation and selection of DNA libraries for directed evolution of proteins also called "protein engineering" are particularly preferred areas of application.

A further preferred use is the production of protein conjugates. Thereby the use is particularly preferred when the protein or protein fragment of the first fusion protein and the protein or protein fragment of the second fusion protein respectively have different requirements for the cellular environment required for correct folding. Thereby the present invention allows the direct fusion of antibodies with marker proteins which would not be correctly folded upon production in bacteria and transport through the Sec-dependent transport pathway and which could, therefore, not be used in standard procedures as marker proteins for marking antibodies. Marker protein antibody fusions the functional expression of which is only enabled by the present invention comprise, for example, fusions of autofluorescent proteins like GFP and immune globulin heavy chains, immune globulin light chains or "single chain antibodies".

The following illustrations and examples are merely provided as an illustration of the invention and not as a limitation to the specific embodiments indicated in the examples. All references comprised in the text are hereby incorporated by reference in their entirety.

## Figures

- Fig. 1 Consensus sequence of Tat-dependent, Sec-dependent, SRP-dependent or YidY-dependent signal sequences wherein X is a random amino acid and # is a hydrophobic amino acid.
- 5 Fig. 2 Tat-dependent TorA-signal peptide, wherein X is a random amino acid and # is a hydrophobic amino acid.
- Fig. 3 Underlying principle of the TLF-system, wherein CT represents the pIII domain, pelB the Sec signal sequence, TSS the Tat signal sequence and POI the presented protein.
- 10 Fig. 4 Restriction map of the plasmid pCD4/GFP24 the nucleic acid sequence of which is depicted in the appendix as SEQ ID NO: 1.
- Fig. 5 Restriction map of the plasmid pCA1/GFP24 the nucleic acid sequence of which is depicted in SEQ ID NO: 2.
- Fig. 6 Restriction map of the plasmid pCN1/GFP24 the nucleic acid of which is depicted in SEQ ID NO: 3.
- 15 Fig. 7 Competitive phage ELISA wherein white bars represent the results with GFP24 presenting phages. GFP24 phages were made with the help of XL-1 blue cells carrying the pCD4/GFP24 plasmid. Grey bars represent the results which were obtained with  $\beta$ -lactamase carrying phages. The  $\beta$ -lactamase presenting phages were made in XL-1 blue cells which carried the plasmid pCD4/BLA.
- 20 Fig. 8 Enzymatic assay of the presentation of  $\beta$ -lactamase on bacteriophages wherein white circles represent the results with GFP24 carrying phages. The GFP24 phages were made with the help of XL-1 blue cells which carry the pCD4/GFP24 plasmid. Black squares represent the results which were obtained with phages carrying the  $\beta$ -lactamase.  $\beta$ -lactamase presenting phages were produced with XL-1 blue cells carrying the pCD4/BLA plasmid. The absorption at 486 nm in relation to the time is shown.
- 25

30

## Examples

### Example 1: Vectors used

pCD4/GFP24 is a cystein display phagimide which is based on the pGP-vector (Paschke M., et al.: (2001) Biotechniques 30: 720-725).

pCAI/GFP24 is a cystein display phagimid which is based on pGF-F100. It can be used for the tet<sup>o-p</sup> controlled expression of proteins as fusion of cFos leucin zipper. The translocation of the cFos-fusion protein into the periplasmatic space is mediated by the TorA leader peptide sequence (Tat-dependent translocation pathway). The tet<sup>o-p</sup> controlled transcript comprises a second cistron which expresses the c-jun::G3Ps fusion protein. The c-Jun::G3Ps is directed into the periplasmatic space through the Sec-dependent translocation pathway. Covalent complexes between the cFos-fusion protein and the cJun::G3PS fusion protein are formed in the periplasmatic space due to the dimerization of cJun and cFos and subsequent formation of cystein bonds between the proteins. The phagimid comprises a GFP24 cassette flanked by SfiI restriction sites at positions 148 and 910 and is positioned between the TorA leader peptide and cFos. This cassette has to be replaced by a protein to be presented.

pCAI/GFP24 is a cystein display phagimid derived from pCD4, which is based on a pGP vector. pDC4/GFP24 is a cystein display phagimid that is based on the pGP vector (Paschke M., et al.: (2001) Biotechniques 30: 720-725). It can be used for the tet<sup>o-p</sup> controlled expression of proteins as fusion with the cFos leucin zipper. The translocation of the cFos fusion protein to the periplasmatic space is mediated by the TorA leader peptide (Tat transport pathway). The tet<sup>o-p</sup> controlled transcript comprises a second cistron, which expresses the c-jun::G3Pss fusion protein (G3Pss comprises amino acids 252 to 406 of the mature gIII proteins of the fd phage). The c-jun::G3Pss is directed towards the periplasmatic space through a Sec-dependent transport pathway (pelB leader peptide). Covalent complexes of cFos fusion protein and c-jun::G3Pss are formed due to the dimerization between cJun and cFos in the periplasmatic space and the subsequent formation of cystein bonds between the proteins (Cramer, R. and Suter M. (1993), supra). The phage display of the proteins, which are fused with cFos can be achieved by so-called helper phage rescue. In contrast to pGP the phagimid pCD4-GFP24 converts chloramphenicol resistance. The resistance gene (CAT) and the tet-repressor (TetR) are under the control of  $\beta$ -lactamase promoter as a bicistronic cassette. The transcript is terminated in a  $\lambda$ -phage terminator. The Tat-TetR cassette is in a reversed position towards the cFos and the cJun fusion cassette. A GFP24 cassette flanked at the positions 148 and 910 by SfiI restrictions sites is positioned between the TorA leader peptide and cFos. This cassette is replaced by the protein to be presented.

pCD4/Bla is a cystein display phagimid derived from pDC4/GFP24 wherein through restriction digest with SfiI the GFP24 fragment has been replaced by the sequence of the



mature TEM1  $\beta$ -lactamase. The inserted lactamase cloning cassette with 5' and 3' terminal SfiI restrictions sites is depicted in SEQ ID NO:4.

#### Example 2: Production of bacterio phages

XL1 blue cells transformed with the respective phagemid were cultivated in 2 TY selection media at 30°C to an OD<sub>600 nm</sub> of 0.5 and then mixed with the helper phage VCSM13 at a Moi = 10-20. The infected culture was cultivated for 30 minutes at 37°C and subsequently mixed with kanamizin at a final concentration of 60 µg/ml. The culture was cultivated at 25°C for 10 minutes. The cells were harvested by centrifugation (4000 x g, 4°C, 5 minutes) and subsequently resuspended in T2Y selection medium comprising 60 µg/ml kanamizin and 0.5 µg/ml tetracycline. The culture was cultivated for 5 hours at 25°C. Subsequently, phage preparation from the cell culture supernatant followed as follows: 40-50 ml cells and cell debris each were separated by centrifugation from phage comprising cell culture supernatant (4°C, 10,000 rpm in an A8-24 rotor for 15 minutes). The supernatant was filtered through a 0.45 µm filter and was mixed with ¼ volume PEG-NaCl solution (20% w/v PEG 8000, 50% w/v NaCl) and incubated on ice over night or at least for 5 hours. The mixture was then centrifuged at 4°C for 15 minutes with 15,000 rpm in an A8.24 rotor. The pellet was resuspended in 2.5 ml ice cold PBS and distributed into 2 ml plastic tubes. Then the supernatant was mixed with ¼ volumes PEG-NaCl solution and incubated on ice for at least one further hour. Then the supernatant was centrifuged at 4°C and 14,000 rpm for 15 minutes. The phage pellet was dissolved in 0.5-1 ml PBS. If necessary the phage solution was filtered through a 0.45 µm filter and then stored at 4°C. For long term storage the phage solution was mixed with 1 volume glycerine and stored at -70°C.

Phage titre was determined using standard methods employing serial dilutions. The titre usually was in the range of 10<sup>12</sup> and 10<sup>13</sup> cfu/ml.

#### Example 3: Presentation of functional GFP24 on phage

GFP24 is a variant of a green fluorescent protein with a circular permutation which further comprises an epitope of the P24 protein of HIV (Höhne, W.E. et al. (1993) Mol. Immunol. 30:1213-21). GFP24 is bound with high affinity by the anti-P24 antibody CB4-1 (Dr. Scholz, Institute for Biochemistry (Universitätsklinikum Charité)). Similar to GFP GFP24 cannot be exported through the Sec-transport pathway. A functional GFP24 protein should resulting from expression of the above described pCD4/GFP24 plasmid should, thus, only lead to the presentation of a functional GFP24, if this part of the protein is not transported into the periplasma by a Sec-dependent transport pathway but rather by a Tat-dependent transport pathway. To detect GFP24 on filamentous phage a phage-ELISA was

carried out as follows: microtitre plates were coated with 10 µm/ml anti-P24 antibody C4-1, washed three times with PBS/Tween 0.1% and incubated for 1 to 2 hours per well with Genosys blocking reagent (Sigma-Genosys Ltd., Cambridge, UK) at room temperature under shaking. Subsequently the microtitre plate was washed three times with PBS/Tween 20 0.1%.

5 Then 50 µl of GFP24 presenting phage with or without P24 peptide was placed in the well of the micro titre plate and then the presence of phage in the microtitre plate was detected with a horseradish peroxidase coupled anti-phage antibody (Seramun Diagnostica GmbH, Dolgenbrodt, Germany). Signal intensity represented the phage bound to CB4-1. pCD4/GFP24 phage was completely replaced from CB4-1 through the P24 peptide while β-  
10 lactamase presenting phages (pCD4/BLA) which were used as a control did not bind to CB4-1. On top of that no unspecific binding to other antibodies or to the blocking reagents could be detected (see Fig. 7).

#### Example 4: Presentation of TEM1-β-lactamase on filamentous phages

TEM-1-β-lactamase is a periplasmatic protein which can confer resistance to  
15 ampicillin by hydrolysis of the lactam ring of the antibiotic ampicillin. TEM-1-β-lactamase is usually exported into the periplasma through a Sec-dependent transport pathway. To show that TEM-1-β-lactamase can also be exported through a Tat-dependent transport pathway the Sec-signal sequence was removed and replaced by the TorA-sequence. The successful presentation of TEM-1-β-lactamase was determined with an enzyme assay described in the  
20 following, the results of which are depicted in Fig. 8. 800 µl PBS pH 7.4 were mixed with 100 µl nitrocefin stock solution (500 µg/ml) and adapted to 25°C. 100 µl phage solution were added. The change of extinction at 486 nm was determined photometrically over 10 min. The change of absorption at 486 nm corresponds to the β-lactamase activity of the phage. While pCD4/GFP24 phages exhibited no β-lactamase activity it was possible to detect a strong β-  
25 lactamase activity for pCD4/BLA.

The nitrocefin stock solution was prepared as follows: 1 mg nitrocefin was dissolved in 100 µl DMSO. The solution was then mixed with 1.9 ml PBS. The solution was stored at -20°C for a maximum of 2 weeks.

### Claims

1. Protein mixture comprising:
  - a) at least a first fusion protein comprising:
    - i) a protein or protein fragment,
    - ii) an interaction domain and
    - iii) a protein translocation sequence which effects that the fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially unfolded state,and
  - b) at least a second fusion protein comprising:
    - i) a protein or protein fragment,
    - ii) an interaction domain and
    - iii) a protein translocation sequence which effects that the fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially folded state,wherein the interaction domain of the first fusion protein can bind to those of the second fusion protein.
2. Protein mixture according to claim 1, wherein the protein or protein fragment of the first fusion protein is an immune globulin heavy chain, an immune globulin light chain, a single chain antibody, a diabody, a receptor, a receptor ligand, an integrin, an intimin, a carbohydrate binding protein, an albumin binding protein or protein A.
3. Protein mixture according to claims 1 or 2, wherein the protein or protein fragment of the second fusion protein is an autofluorescent protein, in particular GFP or a variant thereof, an enzyme, a cofactor-dependent protein, a protein that is encoded by a cDNA derived from a cDNA library or a synthetic protein.
4. Protein mixture according to claim 1, wherein the protein or the protein fragment of the first fusion protein and the protein translocation sequence is a phage coat protein, a

periplasmatic marker enzyme, an intimin, a protein of the outer bacterial membrane or a periplasmatic receptor protein.

5. Protein mixture according to claim 4, wherein the phage coat protein is selected from the M13 phage coat proteins pIII, pVI, pVII, pVIII and pIX.
6. Protein mixture according to claims 1 to 5, wherein the interaction domains of the first and the second fusion protein are each respectively a leucine zipper domain and a leucine zipper domain, a helix-loop-helix-domain and a helix-loop-helix-domain, a calmodulin and a calmodulin binding peptide or a peptid dimer pair of naturally or synthetic origin.
7. Protein mixture according to one of claims 1 to 6, wherein the protein translocation sequence of the first fusion protein is a Sec-dependent, a SRP-dependent, a YidC-dependent sequence or a transport pathway-independent sequence which is integrated into the membrane.
8. Protein mixture according to one of claims 1 to 7, wherein the protein translocation sequence of the second fusion protein is a Tat dependent or  $\Delta$ -ph dependent sequence.
9. Protein mixture according to one of claims 1 to 8, wherein the protein is covalently or non-covalently bound to the second fusion protein.
10. Nucleic acid mixture coding for a protein mixture according to one of claims 1 to 8.
11. Nucleic acid mixture according to claim 10, wherein at least two nucleic acids which code for different fusion proteins are covalently attached to each other.
12. Vector comprising a protein mixture according to one of claims 1 to 9 and/or a nucleic acid mixture according to one of claims 10 or 11.
13. Cell comprising a protein mixture according one of claims 1 to 9, a nucleic acid mixture according to one of claims 10 or 11 and/or a vector according to claim 12.

14. Library comprising at least two protein mixtures according to one of claims 1 to 9, at least two vectors according to claim 12 and/or at least two cells according to claim 13, wherein the proteins or protein fragments of the respective first or the respective second fusion protein are different from each other.
15. Method of identifying a substance, which can bind to a protein mixture according to one of claims 1 to 9, to a vector according to claim 12 or to a cell according to claim 13 comprising the steps:
  - a) contacting at least one potentially binding substance with a protein mixture according to one of claims 1 to 9, a vector according to claim 12, and/or a cell according to claim 13 and
  - b) determining of the binding of the potentially binding substance to the protein mixture, the vector and/or the cell.
16. Method of identifying proteins or protein fragments, which bind to a test substance comprising the following steps:
  - a) contacting at least one test substance with a library according to claim 14 and
  - b) measuring of the respective binding of the test substance to the different protein mixtures, vectors and/or cells of the library.
17. A method according to claim 16 comprising the further steps:
  - a) selecting at least one protein mixture, one vector or one cell on the basis of the measured binding and
  - b) generating a second library wherein the library is generated by modification of the protein or protein fragment comprised in the selected protein mixture, in the selected vector or in the selected cell.
18. Method according to claim 16 comprising the further steps:

- a) selecting at least one protein mixture, one vector or one cell on the basis of the measured binding,
  - b) producing a second library wherein the library is created through the modification of the proteins or protein fragments comprised in the selected protein mixture, in the selected vector or in the selected cell,
  - c) contacting at least one test substance with second library,
  - d) measuring of the respective binding of the test substance to the different protein mixtures, vectors or cells of the second library and
  - e) if the case may be repeating of steps a) to d) until a protein mixture, a vector or a cell is selected which exhibits the desired binding.
19. Method according to one of the claims 15 to 18, wherein in a further step the binding substance of the protein or protein fragment or a variant thereof comprised in the selected protein mixture, in the selected vector or in the selected cell is mixed with a pharmaceutical acceptable carrier and/or auxiliary substance.
20. Kit for the production of a nucleic acid mixture according to claim 10 comprising:
- a) at least a first nucleic acid comprising at least a first restriction site 5' and/or 3' of a nucleic acid coding for a first fusion protein comprising:
    - i) an interaction domain and
    - ii) a protein translocation sequence which effects that the first fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane upon expression in a bacterium in an essentially folded state.
21. Kit according to claim 20 further comprising:

- a) at least a second nucleic acid comprising at least one restriction site 5' and/or 3' of an nucleic acid coding for a second fusion protein comprising:
  - i) an interaction domain and
  - ii) a protein translocation sequence which effects that the second fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially unfolded state,

wherein the interaction domain of the first fusion protein can bind to those of the second fusion protein.

- 22. Use of a cell according to claim 13 for the production of a protein mixture according to one of claims 1 to 9.
- 23. Use of a protein mixture according to one of claims 1 to 9, a vector according to claim 12 and/or a cell according to claim 13 for the production of the library according to claim 14.

### **Summary**

5

The present invention concerns a protein mixture comprising at least a first fusion protein comprising a protein or protein fragment, and an interaction domain and a protein translocation sequence, which effects that the fusion protein upon expression in a bacterium is translocated through the cytoplasmic membrane in an essentially unfolded state and at least a

10 second fusion protein comprising a protein or protein fragment, and an interaction domain and a protein translocation sequence which effects that the fusion protein is translocated through the cytoplasmic membrane upon expression in a bacterium in an essentially folded state, wherein the interaction domain of the first protein can bind to those of the second protein.



## Sequence Listing

&lt;110&gt; Universitätsklinikum Charité

JC20 Rec'd PCT/PTO 06 JUN 2005

&lt;120&gt; Mixture of at least two fusion proteins as well as their production and use

&lt;130&gt; U30038US

&lt;150&gt; DE 10256669.0

&lt;151&gt; 2002-12-04

&lt;160&gt; 4

&lt;170&gt; Word 98, Windows

&lt;210&gt; 1

&lt;211&gt; 4765

&lt;212&gt; DNA

&lt;213&gt; artificial sequence

&lt;220&gt;

&lt;221&gt; pCD4/GFP24 cloning and expression vector

&lt;400&gt; 1

ctagataaga	aggaagaaaa	ataatgaaca	ataacgatct	ctttcaggca	tcacgtcggc	60
gttttctggc	acaactcggc	ggcttaaccg	tgcgcgggat	gctggggccg	tcattgttaa	120
cgccgcgacg	tgcgactgcg	gccagccgg	ccatggcggg	atccgttcaa	ctagcagacc	180
attatcaaca	aaatactcca	attggcgatg	gccctgtcct	tttaccagac	aaccattacc	240
tgtcgacaca	atctgccctt	tcgaaagatc	ccaacgaaaa	gcgtgaccac	atggtccttc	300
ttgagtttgt	aactgctgct	gggatttccg	gtggtggtgg	tgctaccccg	caggacctga	360
acaccatgct	gggtggtggt	ggtagtaaag	gagaagaact	tttcaactgga	gttgtcccaa	420
ttcttggtga	attagatggt	gatgttaatg	ggcacaaatt	ttctgtcagt	ggagaggggtg	480
aaggtgatgc	aacatacggg	aaacttacct	ttaaatttat	ttgcactact	ggaaaactac	540
ctgttccatg	gccaacactt	gtcactactt	tctcttatgg	tggtcaatgc	ttttcccgtt	600
atccggatca	tatgaaacgg	catgactttt	tcaagagtgc	catgcccga	ggttatgtac	660
aggaacgcac	tatatctttc	aaagatgacg	ggaactacaa	gacgcgtgct	gaagtcaagt	720
ttgaaggtga	tacccttggt	aatcgatcgc	agttaaaagg	tattgatttt	aaagaagatg	780
gaaacattct	cggacacaaa	ctcgagtaca	actataactc	acacaatgta	tacatcacgg	840
cagacaaaca	aaagaatgga	atcaaagcta	acttcaaaat	tcgccacaac	attgaagatt	900
cggcctcggg	ggccgcagaa	caaaaactca	tctcagaaga	gaatctgtat	ttccagggcg	960
atgcttgccg	tggcaccgac	accctgcaag	ctgaaaccga	ccagctggaa	gacgagaaat	1020
ccgctctgca	gactgaaatc	gctaacctgc	tgaaagagaa	agagaaactg	gaattcattc	1080
tggctgctca	cggcggttgt	gggctaggct	aataacttaa	gccaaaggag	aaaataaaat	1140
gaaataccta	ttgcctacgc	cagccgctgg	attgttatta	ctcgcggcac	agccggccat	1200
ggcaagcatc	tgcggtggcc	gtatcgctcg	tctggaagaa	aaagttaaaa	ccctgaaagc	1260
tcagaactcc	gaactggctt	ccaccgctaa	catgctgcgt	gaacagggtg	ctcagctgaa	1320
gcagaaagtt	atgaaccacg	gcggttgtgg	tggcggttcc	ctagcgggct	ccggttccgg	1380
tgattttgat	tatgaaaaaa	tggcaaacgc	taataagggg	gctatgaccg	aaaatgccga	1440
tgaaaacgcg	ctacagtctg	acgctaaagg	caaacttgat	tctgtcgcga	ctgattacgg	1500
tgctgctatc	gatggtttca	ttggtgacgt	ttccggcctt	gctaattgga	atggtgctac	1560
tggtgatttt	gctggctcta	attcccaaat	ggctcaagtc	ggtgacggtg	ataattcacc	1620
tttaatgaat	aattttccgc	aatatttacc	ttctttgcct	cagtcgggtg	aatgtcgccc	1680
ttatgtcttt	ggcgctggtg	aacctatga	attttctatt	gattgtgaca	aaataaaact	1740
attccgtggt	gtctttgctg	ttcttttata	tggttgcacc	tttatgtatg	tattttcgac	1800
gtttgctaac	atactgcgta	ataaggagtc	ttaataagct	tgacctgtga	agtgaaaaat	1860
ggcgcacatt	gtgcgacatt	ttttttgtct	gccgtttacc	gctactgcgt	cacggatctc	1920
cacgcgccct	gtagcggcgc	attaagcgcg	gcgggtgtgg	tggttacgcg	cagcgtgacc	1980
gctacacttg	ccagcgcctt	agcgcgcctt	cctttcgctt	tcttcccttc	ctttctcgcc	2040
acgttcgccc	gctttccccc	tcaagctcta	aatcgggggc	tccctttagg	gttccgattt	2100
agtgttttac	ggcacctcga	ccccaaaaaa	cttgattagg	gtgatggttc	acgtagtggg	2160
ccatcgccct	gatagacggt	ttttcgccct	ttgacgttgg	agtccacgtt	ctttaatagt	2220

ggactcttgt	tccaaactgg	aacaacactc	aaccctatct	cggtctatct	ttttgattta	2280
taagggattt	tgccgatttc	ggcctattgg	ttaaaaaatg	agctgattta	acaaaaattt	2340
aacgcgcatg	ctaacaaaat	attaaaaaac	gcccggcggc	aaccgagcgt	taatagttaa	2400
gttaccatca	cggaaaaagg	ttatgctgct	tttaagaccc	actttcacat	ttaggttgtt	2460
tttctaattc	gcatatgata	aattcaaggc	cgaataagaa	ggctggctct	gcaccttggt	2520
gatcaaataa	ttcgatagct	tgtcgtaata	atggcgccat	actatcagta	gtaggtgttt	2580
ccctttcttc	tttagcgact	tgatgctctt	gatcttccaa	tacgcaacct	aaagtaaaat	2640
gccccactgc	gctgagtgc	tataatgcat	tctctagtga	aaaaccttgt	tggcataaaa	2700
aggctaattg	atthttcgaga	gtttcatact	gtttttctgt	aggccgtgta	cctaaatgta	2760
cttttgctcc	atcgcgatga	cttagtaaaag	cacatctaaa	acttttagcg	ttattacgta	2820
aaaaactctg	ccagctttcc	ccttctaaag	ggcaaaaagt	agtatggtgc	ctatctaaca	2880
tctcaatggc	taaggcgctg	agcaaaagcc	gcttatthtt	tacatgccaa	tacaatgtag	2940
gctgctctac	acctagcttc	tgggaggttt	tacgggttgt	taaaccttcg	attccgacct	3000
cattaagcag	ctctaattgc	ctgttaatac	ctttactttt	atctaaacga	gacatcatta	3060
attcctatta	cgccccgccc	tgccactcat	cgcagtactg	ttgtaattca	ttagcatttc	3120
tgccgacatg	gaagccatca	caaacggcat	gatgaacctg	aatcgccagc	ggcatcagca	3180
ccttgctcgc	ttgctgataa	tatttgccca	tagtgaaaac	gggggagaa	aagttgtcca	3240
tattggccac	gtttaaatca	aaactgggtga	aactcaccca	gggattggct	gagacgaaaa	3300
acataattct	aataaaacct	ttagggaaat	aggccaggtt	ttcaccgtaa	cacgccacat	3360
cttgcgataa	tatgtgtaga	aactgcccga	aatcgctcgt	gtattcactc	cagagcgatg	3420
aaaacgtttc	agtttgctca	tggaaaaagg	tgtaacaagg	gtgaacacta	tcccatatca	3480
ccagctcacc	gtctttcatt	gccatacggg	attccggatg	agcattcatc	aggcgggcaa	3540
gaatgtgaat	aaaggccgga	taaaacttgt	gcttatthtt	ctttacggtc	tttaaaagg	3600
ccgtaatatc	cagctgaacg	gtctgggtat	aggtacattg	agcaactgac	tgaatgcct	3660
caaaatgttc	tttacgatgc	cattgggata	tatcaacggg	ggtatatcca	gtgatttttt	3720
tctccatact	cttcctthtt	caatattatt	gaagcattta	tcagggttat	tgtctcatga	3780
gcggatacat	atthgaatgt	atthtagaaa	ataaacaaat	agggttccg	cgcacatttc	3840
cccgaaaagt	gccacctgaa	attgtaagcg	ttactagttt	aaaaggatct	agggtgaagat	3900
cctthtttgt	aatctcatga	ccaaaatccc	ttaacgtgag	ttttcgthtc	actgagcgtc	3960
agaccccgta	gaaaagatca	aaggatcttc	ttgagatcct	ttthttctgc	gcgtaactct	4020
ctgcttgcaa	acaaaaaaac	caccgctacc	agcgtgtggt	tgthtgccgg	atcaagagct	4080
accaactctt	tttccgaagg	taactggctt	cagcagagcg	cagataccaa	atactgtcct	4140
tctagtgtag	ccgtagttag	gccaccactt	caagaactct	gtagcaccgc	ctacatacct	4200
cgctctgcta	atcctgttac	cagtggctgc	tgccagtggc	gataagtcgt	gtcttaccgg	4260
gttgactca	agacgatagt	taccggataa	ggcgagcg	tcgggctgaa	cggggggttc	4320
gtgcacacag	cccagcttgg	agcgaacgac	ctacaccgaa	ctgagatacc	tacagcgatg	4380
gctatgagaa	agcgccacgc	ttcccgaagg	gagaaaggcg	gacaggatct	cggtaagcgg	4440
cagggtcgga	acaggagagc	gcacgaggga	gcttccaggg	ggaaacgcct	ggtatcttta	4500
tagtctgttc	gggtttcgcc	acctctgact	tgagcgtcga	ttthttgtat	gctcgtcagg	4560
ggggcgagc	gctatgaaaa	acgccagcaa	cgcggctctt	ttacggthtc	tggcctthttg	4620
ctggccttht	gctcacatga	cccgacacca	tcgaatggcc	agatgattaa	ttcctaattt	4680
ttgttgacac	tctatcattg	atagagttat	tttaccactc	cctatcagtg	atagagaaaa	4740
gtgaaatgaa	tagttcgaca	aaaat				4765

<210> 2

<211> 4971

<212> DNA

<213> artificial sequence

<220>

<221> pCA1/GFP24 cloning and expression vector

<400> 2

ctagataaga	aggaagaaaa	ataatgaaca	ataacgatct	ctttcaggca	tcacgtcggc	60
gtthttctggc	acaactcggc	ggcttaaccg	tcgccgggat	gctggggccg	tcattgttaa	120
gcgccgacag	tgcgactgcg	gcccagccgg	ccatggcggg	atccgttcaa	ctagcagacc	180
attatcaaca	aaatactcca	attggcgatg	gccctgtcct	tttaccagac	aaccattacc	240
tgtcgacaca	atctgccctt	tcgaaagatc	ccaacgaaaa	gcgtgaccac	atggtccttc	300
ttgagthttgt	aactgctgct	gggattthccg	gtgggtgggtg	tgctaccccg	caggacctga	360
acaccatgct	gggtgggtgg	ggtagttaaag	gagaagaact	tttacttgga	gttgtcccaa	420
ttcttgthtga	attagatggg	gatgttaaatg	ggcacaattt	ttctgtcagt	ggagaggggtg	480
aagggtgatgc	aacatacggg	aaacttacct	ttaaatttat	ttgcactact	ggaaaactac	540

ctgttccatg	gccaacactt	gtcactactt	tctcttatgg	tgttcaatgc	ttttcccggt	600
atccggatca	tatgaaacgg	catgactttt	tcaagagtgc	catgcccga	ggttatgtac	660
aggaacgcac	tatatctttc	aaagatgacg	ggaactacaa	gacgcgtgct	gaagtcaagt	720
ttgaagggtga	tacccttggt	aatcgatatcg	agttaaaagg	tattgatttt	aaagaagatg	780
gaaacattct	cggacacaaa	ctcgagtaca	actataactc	acacaatgta	tacatcacgg	840
cagacaaaca	aaagaatgga	atcaaagcta	acttcaaaat	tcgccacaac	attgaagatt	900
cggcctcggg	ggccgcagaa	caaaaactca	tctcagaaga	gaatctgtat	ttccagggcg	960
ggcccaaacc	ttccaccccc	cctgggttctt	caggcgcctg	cgggtggcctg	accgacaccc	1020
tgcaagctga	aaccgaccag	ctggaagacg	agaaatccgc	tctgcagact	gaaatcgcta	1080
acctgctgaa	agagaaagag	aaactggaat	tcattctggc	tgctcacggc	ggttgttaat	1140
aacttaagcc	aaggaggaaa	ataaaatgaa	atacctattg	cctacggcag	ccgctggatt	1200
gttattactc	gctgcccac	cagcgatggc	cgcacagggt	aaactgctcg	agagcgcttg	1260
cggtgccgt	atcgctcgtc	tggagaaaaa	agttaaaacc	ctgaaagctc	agaactccga	1320
actggcttcc	accgctaaca	tgctgcgtga	acaggttgct	cagctgaagc	agaaagttat	1380
gaaccacggc	ggttgtgcta	gcggtggcgg	ctccggttcc	ggtgattttg	attatgaaaa	1440
aatggcaaac	gctaataagg	gggctatgac	cgaaaatgcc	gatgaaaacg	cgctacagtc	1500
tgacgctaaa	ggcaaacttg	attctgtcgc	tactgattac	ggtgctgcta	tcgatggttt	1560
cattggtgac	gtttccggcc	ttgctaattg	taatggtgct	actggtgatt	ttgctggctc	1620
taattcccaa	atggctcaag	tcggtgacgg	tgataattca	cctttaatga	ataatttccg	1680
tcaatattta	ccttctttgc	ctcagtcggt	tgaatgtcgc	ccttatgtct	ttggcgctgg	1740
taaaccatat	gaattttcta	ttgattgtga	caaaataaac	ttattccgtg	gtgtctttgc	1800
gtttctttta	tattgtgcca	cctttatgta	tgtattttcg	acgtttgcta	acatactcg	1860
taataaggag	tcttaataag	cttgacctgt	gaagtgaaaa	atggcgca	ttgtgcgaca	1920
ttttttttgt	ctgccgttta	ccgctactgc	gtcacggatc	tcacgcgc	ctgtagcggc	1980
gcattaagcg	cggcggtgt	ggtggttacg	cgcagcgtga	ccgctacact	tgccagcgcc	2040
ctagcgcgcg	ctcctttcgc	tttcttccct	tcctttctcg	ccacgttcgc	cggctttccc	2100
cgtcaagctc	taaatcgggg	gctcccttta	gggttccgat	ttagtgttt	acggcacctc	2160
gaccccaaaa	aacttgatta	gggtgatggt	tcacgtagt	ggccatcgcc	ctgatagacg	2220
gtttttcgcc	ctttgacgtt	ggagtccacg	ttctttaata	gtggactctt	gttccaaaact	2280
ggaacaacac	tcaaccctat	ctcgggtctat	tcttttgatt	tataagggat	tttgccgatt	2340
tcggcctatt	ggttaaaaaa	tgagctgatt	taacaaaaat	ttaacgcgaa	ttttaacaaa	2400
atattaacgc	ttacaatttc	aggtggcact	tttcgggaa	atgtgcgcgg	aacctctatt	2460
tgtttatttt	tctaaatata	ttcaaatatg	tatccgctca	tgagacaata	accttgataa	2520
atgcttcaat	aatattgaaa	aaggaagagt	atgagtattc	aacattttcg	tgtgcgcctt	2580
attccctttt	ttgcggcatt	ttgccttcct	gtttttgctc	accagaaac	gctggtgaaa	2640
gtaaaagatg	ctgaagatca	ggtgggtgca	cgagtgggtt	acatcgaact	ggatctcaac	2700
agcggtaaga	tccttgagag	ttttcgcccc	gaagaacgtt	ttccaatgat	gagcactttt	2760
aaagtctctg	tatgtggcgc	ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcgg	2820
cgccgcatac	actatttcta	gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	2880
cttacggatg	gcatgacagt	aagagaatta	tgagtgctg	ccataaccat	gagtataaac	2940
actgcggcca	acttacttct	gacaacgatac	ggaggaccga	aggagctaac	cgtgtttttg	3000
cacaacatgg	gggatcatgt	aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	3060
ataccaaacg	acgagcgtga	caccacgatg	cctgtagcaa	tggaacaac	gttgcgcaaa	3120
ctattaactg	gcgaactact	tactctagct	tcccggaac	aattgataga	ctggatggag	3180
gcggataaag	ttgcaggacc	acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	3240
gataaatctg	gagccggtga	gcgtggctct	cgcggtatca	ttgcagcact	ggggccagat	3300
ggtaaagcct	cccgtatcgt	agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	3360
cgaatatagac	agatcgctga	gataggtgcc	tactgatta	agcattggta	ggaattaatg	3420
atgtctcggt	tagataaaaa	taaagtgatt	aacagcgcat	tagagctgct	taatgaggtc	3480
ggaatcgaag	gtttaacaac	ccgtaaactc	gccagaagc	taggtgtaga	gcagcctaca	3540
ttgtattggc	atgtaaaaaa	taagcgggct	ttgctcgacg	ccttagccat	tgagatgtta	3600
gataggcacc	atactcactt	ttgcctttta	gaaggggaaa	gctggcaaga	ttttttacgt	3660
aataacgcta	aaagtttttag	atgtgcttta	ctaagtcata	gcgatggagc	aaaagtacat	3720
ttaggtacac	ggcctacaga	aaaacagtat	gaaactctcg	aaaatcaatt	agccttttta	3780
tgccaacaag	gttttttact	agagaatgca	ttatatgcac	tcagcgcagt	ggggcatttt	3840
actttaggtt	gcgtattgga	agatcaagag	catcaagtcg	ctaaagaaga	aagggaacaa	3900
cctactactg	atagtatgcc	gccattatta	cgacaagcta	tcgaattatt	tgatcaccaa	3960
ggtgcagagc	cagccttctt	attcggcctt	gaattgatca	tatgcggatt	agaaaaacaa	4020
cttaaatgag	aaagtgggtc	ttaaaagcag	cataaccttt	ttccgtgatg	gtaacttcac	4080
tagtttaaaa	ggatctagg	gaagatcctt	tttgataatc	tcatgaccaa	aatcccttaa	4140
cgtgagtttt	cgttccaactg	agcgtcagac	cccgtagaaa	agatcaaagg	atcttcttga	4200
gatccttttt	ttctgcgcgt	aatctgctgc	ttgcaaacaa	aaaaaccacc	gctaccagcg	4260
gtggtttgtt	tgccggatca	agagctacca	actctttttc	cgaaggtaac	tggcttcagc	4320

agagcgcaga	taccaaatac	tgtcctttcta	gtgtagccgt	agttaggcca	ccacttcaag	4380
aactctgtag	caccgcctac	atacctcgct	ctgctaatacc	tgttaccagt	ggctgctgcc	4440
agtggcgata	agtcgtgtct	taccgggttg	gactcaagac	gatagttacc	ggataaggcg	4500
cagcggtcgg	gctgaacggg	gggttcgtgc	acacagccca	gcttggaagc	aacgacctac	4560
accgaactga	gatacctaca	gcgtgagcta	tgagaaagcg	ccacgcttcc	cgaagggaga	4620
aaggcggaca	ggtatccggt	aagcggcagg	gtcggaaacag	gagagcgcac	gagggagctt	4680
ccagggggaa	acgcctggta	tctttatagt	cctgtcgggt	ttcgccacct	ctgacttgag	4740
cgtcgatttt	tgtgatgctc	gtcagggggg	cggagcctat	ggaaaaacgc	cagcaacgcg	4800
gcctttttac	ggttcctggc	cttttgctgg	ccttttgctc	acatgaccgc	acaccatcga	4860
atggccagat	gattaattcc	taatttttgt	tgacactcta	tcattgatag	agttatttta	4920
ccactcccta	tcagtgatag	agaaaagtga	aatgaatagt	tcgacaaaaa	t	4971

<210> 3  
 <211> 4765  
 <212> DNA  
 <213> artificial sequence

<220>  
 <221> pCN1/GFP24 cloning and expression vector

<400> 3

ctagataaga	aggaagaaaa	ataatgaaca	ataacgatct	ctttcaggca	tcacgtcggc	60
gtttttctggc	acaactcggc	ggcttaaccg	tcgccgggat	gctggggccg	tcattgttaa	120
cgccgcgacg	tgcgactgcg	gccagccgg	ccatggcggg	atccgttcaa	ctagcagacc	180
attatcaaca	aaatactcca	attggcgatg	gccctgtcct	tttaccagac	aaccattacc	240
tgtcgacaca	atctgccctt	tcgaaagatc	ccaacgaaaa	gcgtgaccac	atggtccttc	300
ttgagtttgt	aactgctgct	gggattttccg	gtgggtgggtg	tgctaccccg	caggacctga	360
acaccatgct	gggtgggtgt	ggtagtaaag	gagaagaact	tttacttgga	gttgtcccaa	420
tcttggtga	attagatggt	gatgttaatg	ggcaciaaatt	ttctgtcagt	ggagagggtg	480
aaggtgatgc	aacatacggg	aaacttacc	ttaaatttat	ttgcactact	ggaaaactac	540
ctgttccatg	gccaacactt	gtcactactt	tctcttatgg	tggtcaatgc	ttttcccggt	600
atccggatca	tatgaaacgg	catgactttt	tcaagagtgc	catgcccgaa	ggttatgtac	660
aggaacgcac	tatatctttc	aaagatgacg	ggaactacaa	gacgcgtgct	gaagtcaagt	720
ttgaagggtga	tacccttggt	aatcgtatcg	agttaaaagg	tattgatttt	aaagaagatg	780
gaaacattct	cggacacaaa	ctcgagtaca	actataactc	acacaatgta	tacatcacgg	840
cagacaaaca	aaagaatgga	atcaaagcta	acttcaaaat	tcgccacaa	attgaagatt	900
cggcctcggg	ggccgcagaa	caaaaactca	tctcagaaga	gaatctgtat	ttccagggcg	960
atgcttgccg	tggcaccgac	accctgcaag	ctgaaaccga	ccagctggaa	gacgagaaat	1020
ccgctctgca	gactgaaatc	gctaaccctg	tgaaagagaa	agagaaactg	gaattcattc	1080
tggctgtctca	ggcggttgt	gggctaggct	aataacttaa	gccaaggagg	aaaataaaat	1140
gaaataccta	ttgcctacgg	cagccgctgg	attgttatta	ctcgcggcac	agccggccat	1200
ggcaagcatc	tgcggtggcc	gtatcgctcg	tctggaagaa	aaagttaaaa	ccctgaaagc	1260
tcagaactcc	gaactggctt	ccaccgctaa	catgctgcgt	gaacagggtg	ctcagctgaa	1320
gcagaaaagt	atgaaccacg	gcggttggtg	tggcggttcc	ctagcgggct	ccggttcccg	1380
tgattttgat	tatgaaaaaa	tggcaaacgc	taataagggg	gctatgaccg	aaaatgccga	1440
tgaaaacgcg	ctacagtctg	acgctaaagg	caaacttgat	tctgtcgcta	ctgattacgg	1500
tgctgctatc	gatggtttca	ttggtgacgt	ttccggcctt	gctaattggt	atggtgctac	1560
tggtgatttt	gctggctcta	attcccaaat	ggctcaagtc	ggtgacggtg	ataattcacc	1620
tttaatgaat	aattttccgtc	aatatattacc	ttctttgcct	cagtcggttg	aatgtcgcgc	1680
ttatgtcttt	ggcgctggta	aaccatatga	attttctatt	gattgtgaca	aaataaactt	1740
attccgtggg	gtctttgcgt	ttcttttata	tggtgccacc	tttatgtatg	tattttcgac	1800
gtttgctaac	atactgcgta	ataaggagtc	ttaataagct	tgacctgtga	agtgaaaaat	1860
ggcgcacatt	gtgcgacatt	ttttttgtct	gccgtttacc	gctactgcgt	cacggatctc	1920
cacgcgccct	gtagcggcgc	attaagcgcg	gcgggtgtgg	tggttacgcg	cagcgtgacc	1980
gctacacttg	ccagcgccct	agcgcccgct	cctttcgctt	tcttcccttc	ctttctcgcc	2040
acgttcgcgc	gctttccccc	tcaagctcta	aatcgggggc	tccctttagg	gttccgattt	2100
agtgctttac	ggcacctcga	ccccaaaaaa	ccttgattagg	gtgatgggtc	acgtagtggg	2160
ccatcgccct	gatagacggg	ttttcgccct	ttgacgttgg	agtccacggt	ctttaatagt	2220
ggactcttgt	tccaaactgg	aacaacactc	aaccctatct	cgggtctattc	ttttgattta	2280
taagggattt	tgccgatttc	ggcctatttg	ttaaaaaatg	agctgattta	acaaaaattt	2340
aacgcgcgatg	caacgccttac	aatttcagggt	ggcacttttc	ggggaaatgt	gcgcggaacc	2400

cctatttgtt	tatttttcta	aatacattca	aatatgtatc	cgctcatgag	acaataaccc	2460
tgataaatgc	ttcaataata	ttgaaaaagg	aagagtatgg	agaaaaaaat	cactggatat	2520
accaccgttg	atatatccca	atggcatcgt	aaagaacatt	ttgaggcatt	tcagtcagtt	2580
gctcaatgta	cctataacca	gaccgttcag	ctggatatta	cggccttttt	aaagaccgta	2640
aagaaaaata	agcacaagtt	ttatccggcc	tttattcaca	ttcttgcccg	cctgatgaat	2700
gctcatccgg	aattccgtat	ggcaatgaaa	gacgggtgagc	tggtgatatg	ggatagtgtt	2760
cacccttgtt	acaccgtttt	ccatgagcaa	actgaaacgt	tttcatcgct	ctggagtga	2820
taccacgacg	atttccggca	gtttctacac	atatattcgc	aagatgtggc	gtgttacggt	2880
gaaaacctgg	cctattttccc	taaagggttt	attgagaata	tgtttttcgt	ctcagccaat	2940
ccctgggtga	gtttcaccag	ttttgattta	aacgtggcca	atatggacaa	cttcttcgcc	3000
cccgttttca	ctatgggcaa	atattatacg	caaggcgaca	aggtgctgat	gccgctggcg	3060
attcaggttc	atcatgccgt	ttgtgatggc	ttccatgtcg	gcagaatgct	taatgaatta	3120
caacagtact	gcgatgagt	gcagggcg	gcgtaatagg	aattaatgat	gtctcgttta	3180
gataaaagta	aagtgtat	caagcgatta	gagctgctta	atgaggtcgg	aatcggaagt	3240
ttaacaaccc	gtaaaactgc	ccagaagcta	ggtgtagagc	agcctacatt	gtattggcat	3300
gtaaaaaata	agcgggcttt	gctcgacgcc	ttagccattg	agatgttaga	taggcaccat	3360
actcactttt	gcccttttaga	aggggaaagc	tggcaagatt	ttttacgtaa	taacgctaaa	3420
agtttttagat	gtgctttact	aagtcacgc	gatggagcaa	aagtacattt	aggtacacgg	3480
cctacagaaa	aacagtatga	aactctcgaa	aatcaattag	ccttttttatg	ccaacaaggt	3540
ttttcactag	agaatgcatt	atatgcactc	agcgcagtgg	ggcattttac	tttaggttgc	3600
gtattggaag	atcaagagca	tcaagtcgct	aaagaagaaa	gggaaacacc	tactactgat	3660
agtatgccgc	cattattacg	acaagctatc	gaattatttg	atcaccaagg	tgcagagcca	3720
gccttcttat	tcggccttga	attgatcata	tgcggttag	aaaaacaact	taaatgtgaa	3780
agtgggtctt	aaaagcagca	taaccttttt	ccgtgatgg	aacttacta	ttaacgctcg	3840
gttgccgccc	ggcggttttt	aatattttgt	taactagttt	aaaaggatct	aggtgaagat	3900
cctttttgat	aatctcatga	ccaaaatccc	ttaacgtgag	ttttcgttcc	actgagcgtc	3960
agaccccgta	gaaaagatca	aaggatcttc	ttgagatcct	ttttttctgc	gcgtaatctg	4020
ctgcttgcaa	acaaaaaaac	caccgctacc	agcggtggtt	tgtttgcccg	atcaagagct	4080
accaactctt	tttccgaagg	taactggctt	cagcagagcg	cagataccaa	atactgtcct	4140
tctagtgtag	ccgtagttag	gccaccactt	caagaactct	gtagcaccgc	ctacatacct	4200
cgctctgcta	atcctgttac	cagtggctgc	tgccagtggc	gataagtcgt	gtcttaccgg	4260
gttggactca	agacgatagt	taccggataa	ggcgagcgg	tcgggctgaa	cggggggttc	4320
gtgcacacag	cccagcttgg	agcgaacgac	ctacaccgaa	ctgagatacc	tacagcgtga	4380
gctatgagaa	agcgccacgc	ttcccgaagg	gagaaaggcg	gacaggtatc	cggtaagcgg	4440
cagggtcgga	acaggagagc	gcacgaggga	gcttccaggg	ggaaacgcct	ggtatcttta	4500
tagtcctgtc	gggtttcgcc	acctctgact	tgagcgtcga	tttttgtgat	gctcgtcagg	4560
ggggcgagc	ctatggaaaa	acgccagcaa	cgcggccttt	ttacggttcc	tggccttttg	4620
ctggcctttt	gtcacatga	cccgaaccca	tcgaatggcc	agatgattaa	ttcctaattt	4680
ttgttgacac	tctatcattg	atagagttat	tttaccactc	cctatcagtg	atagagaaaa	4740
gtgaaatgaa	tagtttcgaca	aaaat				4765

<210> 4  
 <211> 823  
 <212> DNA  
 <213> artificial

<220>  
 <221> mature TEM-1  $\beta$ -lactamase cloning cassette

<400> 4

ggcccagccg	gccatggctc	accagaaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	60
gttgggtgca	cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	120
ttttcgcccc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	280
ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcgg	cgccgcatac	actattctca	240
gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	300
aagagaatta	tgcagtgtcg	ccataaccat	gagtataaac	actgcccga	acttacttct	360
gacaacgac	ggagagccga	aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	420
aactcgcctt	gactcgttgg	aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	480
caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	540
tactctagct	tcccggcaac	aattgataga	ctggatggag	gcggataaag	ttgcaggacc	600

acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccggtga	660
gcgtggctct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt	720
agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	780
gatagggtgcc	tcactgatta	agcattggtc	ggcctcgggg	gcc		823

← about 26-58 →

⊕ charged terminus      central hydrophobic      recognition sequence

A X A

M N N N D L F Q A R R R F A Q L G G L T  
ATGAACA ATAACGATCT CTTTCAGGCA TCACGTCCTTGGC ACAACTCGGC  
TACTTGT TATTGCTAGA GAAAGTCCGT AGTGCAGACACCG TGTTGAGCCG

V A G M L G P S L L T R R A T A  
GTCGGGGAT GCTGGGGCGATTGTAA CGCCGCGACG  
CACGGCCTA CGACCCCGGC AGTAACAATCGCTGC

The diagram shows a protein structure with a central hydrophobic core and a leader peptidase recognition sequence. The central hydrophobic core is represented by a thick black line. The leader peptidase recognition sequence is represented by a box containing the sequence A X A.



Fig. 3

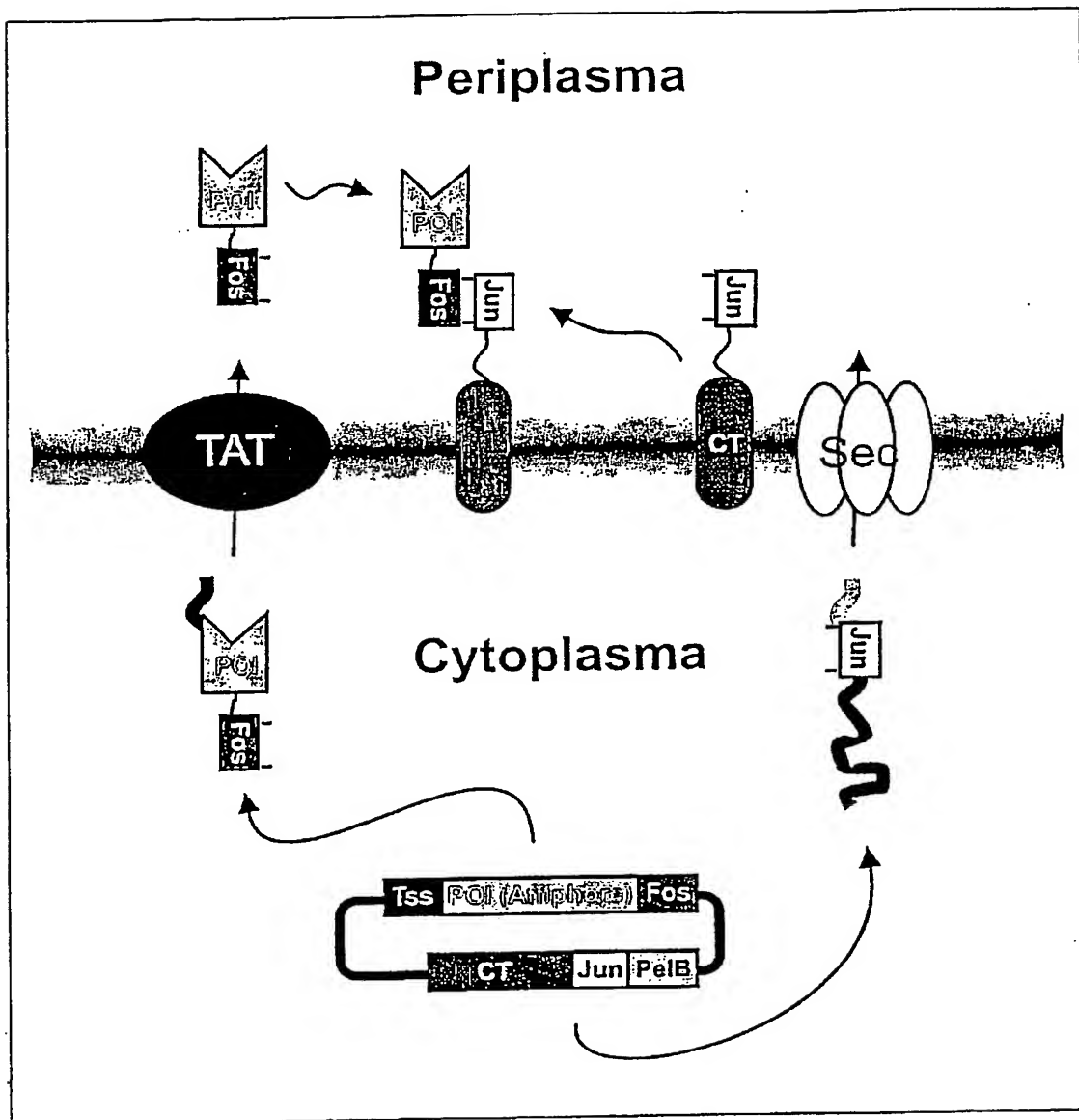


Fig. 4

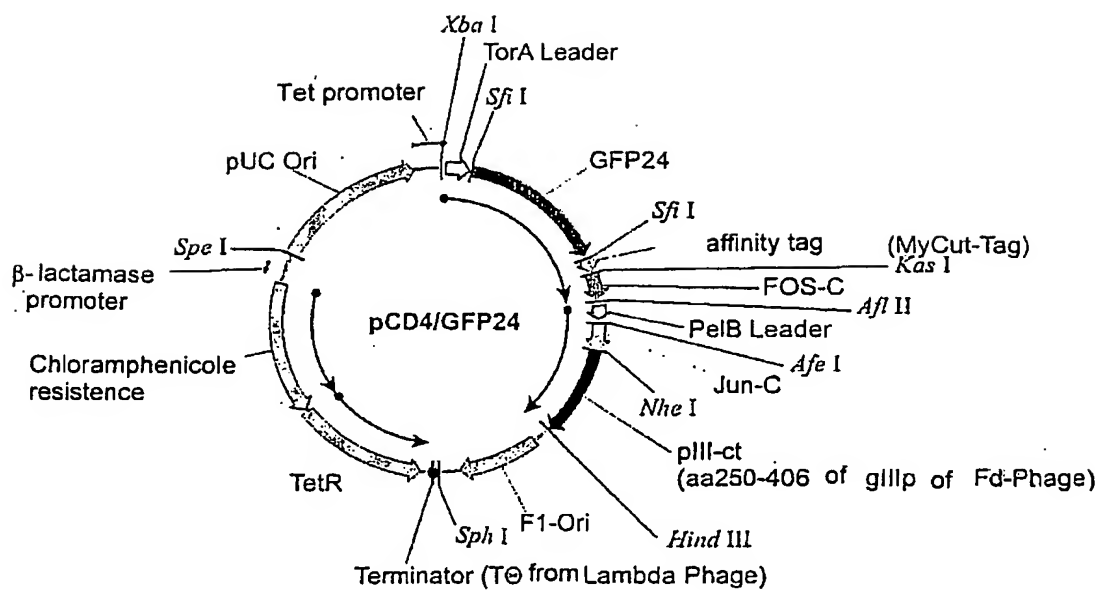


Fig. 5

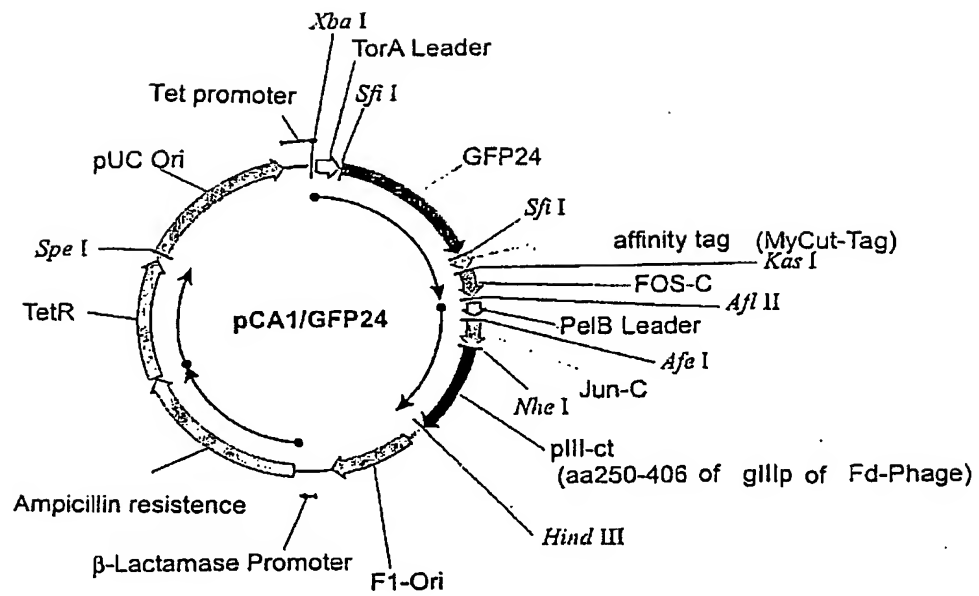


Fig. 6

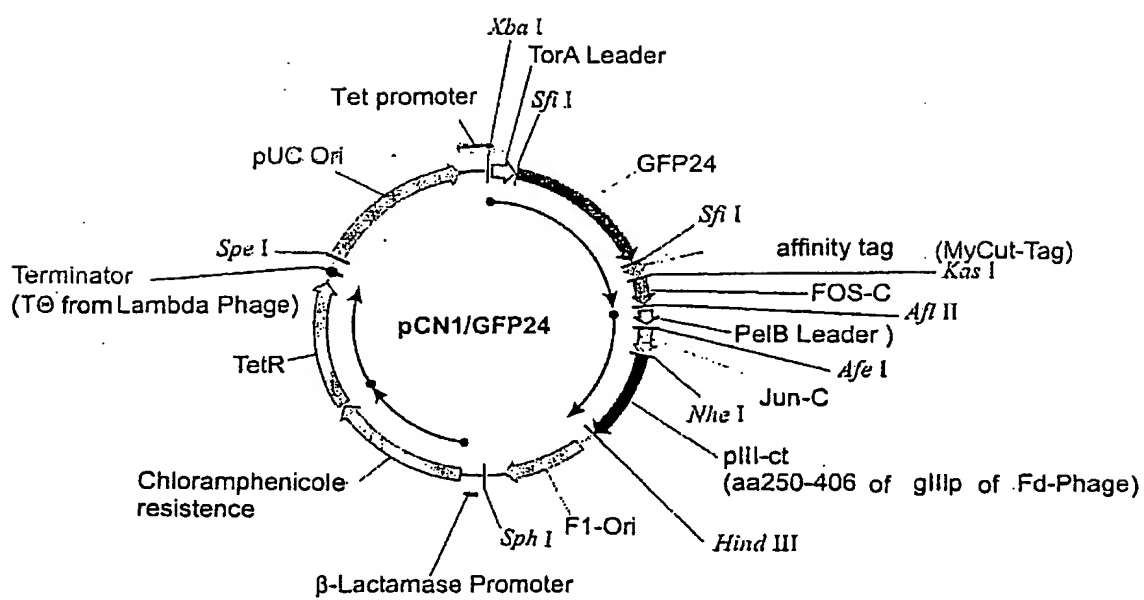
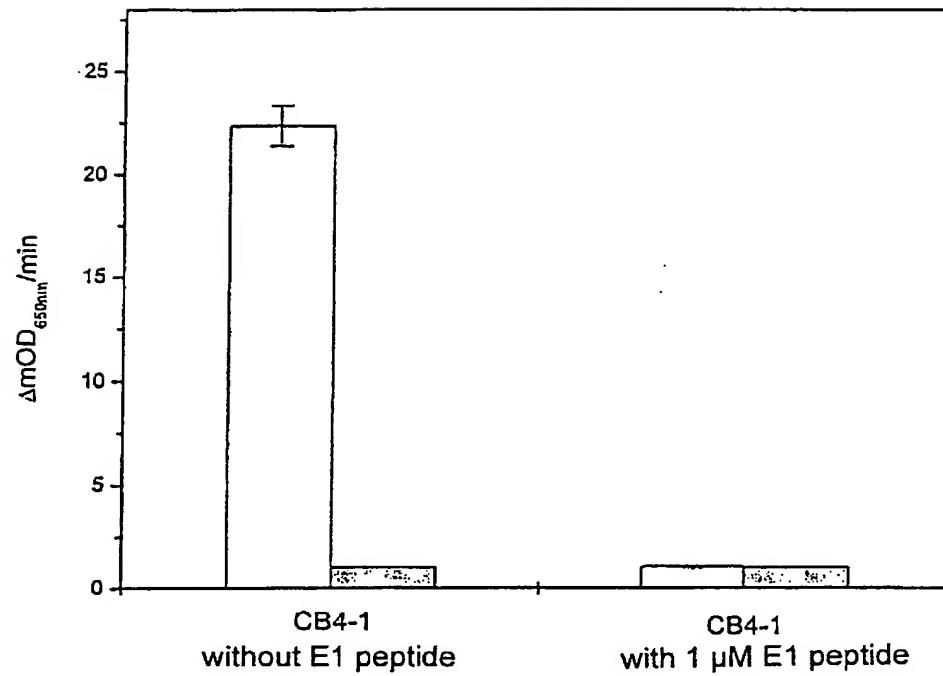


Fig. 7



8/8

Fig. 8

